Proc. Indian Acad. Sci. (Math. Sci.) Vol. 116, No. 2, May 2003, pp. 147–160.
Printed in India

Non-Euclidean visibility problems

FERNANDO CHAMIZO

Departamento de Matemáticas, Facultad de Ciencias, Universidad Autónoma de Madrid, 28049 Madrid, Spain

MS received 14 November 2005

Abstract. We consider the analog of visibility problems in hyperbolic plane (represented by Poincaré half-plane model $\mathbb{H}$), replacing the standard lattice $\mathbb{Z}\times\mathbb{Z}$ by the orbit $z=i$ under the full modular group $\mathrm{SL}_2(\mathbb{Z})$. We prove a visibility criterion and study orchard problem and the cardinality of visible points in large circles.

Keywords. Modular group; hyperbolic plane; Poincaré half-plane model.

1. Introduction

Consider in $\mathbb{R}^2$ the standard lattice $L=\mathbb{Z}\times\mathbb{Z}$ and the origin $(0,0)$. A point $(m,n)\in L-\{(0,0)\}$ is said to be *visible* if the segment connecting the origin and (m,n) does not contain any other lattice points.

Visibility problems have been studied since a century. Perhaps the most celebrated problems are the visible version of Gauss circle problem and the so-called orchard problem (see other problems in [5]). In both of these problems one considers visible lattice points in a large circle. The first problem consists of approximating the cardinality of this set of points. It turns out that improvements on the trivial bounds of the error term are related to Riemann Hypothesis (see [12]). In the orchard problem the visible points are considered to be thick and it is asked the minimal thickness such that all exterior points are eclipsed. In the formulation included in p. 150 of [14], “How thick must the trunks of the trees in a regularly spaced circular orchard grow if they are to block completely the view from the center?”. In contrast with the previous problem, orchard problem can be considered as solved in a wide sense (see [2]) by elementary methods.

In this paper we deal with the hyperbolic analog of visibility problems. Namely, we consider Poincaré’s plane $\mathbb{H}$, i.e., the upper half plane endowed with the metric

$$\mathrm{d}s^2=y^{-2}\,\mathrm{d}x^2+y^{-2}\,\mathrm{d}y^2, \tag{1}$$

the origin $i\in\mathbb{H}$ and $\mathscr{L}$ to be the orbit of $z=i$ under the full modular group $\mathrm{SL}_2(\mathbb{Z})$ (note that in the Euclidean case the lattice $\mathbb{Z}\times\mathbb{Z}$ is the orbit of the origin under the discrete group formed by all integral translations). We say that $z\in\mathscr{L}$, $z\neq i$ is *visible* if the arc of geodesic connecting i and z does not contain any other point in $\mathscr{L}$.

One cannot draw a parallel between the study of visibility problems in a hyperbolic case and the Euclidean case due to the following algebraic and geometric facts: Firstly, in the Euclidean case the group of integral translations is Abelian, but in the hyperbolic case the underlying group $\mathrm{SL}_2(\mathbb{Z})$ is not. Secondly, the Euclidean isoperimetric inequality

$4\pi A \le l^2$, which is sharp for circles, is qualitatively different from its hyperbolic analog $4\pi A + A^2 \le l^2$, for large areas (p. 11 of [10]).

We shall structure each of the following sections stating Euclidean results first and then their hyperbolic counterparts; this will ease the comparison between both settings. After studying the symmetries and some other preliminary topics in §2, we give in §3 a hyperbolic criterion for the visibility of a point and investigate the structure of 'lattice' points in rays. Visible Gauss circle problem and orchard problem are discussed in §§4 and 5. Finally, in §6 we show some numerical data to illustrate our results.

As an aside, we want to point out that although our main motivation is number theoretical, visibility problems have called the attention of some physicists (e.g., [1] and [3]) and it is plausible that the change of the geometry could be meaningful in some applications. For instance, Olbers' paradox (which even after two centuries still motivates some research and controversy [15]) in an idealized sharper form taking into account occultation could lead to considerations about visible points in some non-Euclidean space.

Notation: As usual, we shall represent with the same symbols a matrix in $\mathrm{SL}_2(\mathbb{Z})$ and its associated fractional linear transformation acting on $\mathbb{H}$:

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \longleftrightarrow \gamma(z) = \frac{az+b}{cz+d}.$$

(Of course there is some ambiguity that would be avoided considering $\mathrm{PSL}_2(\mathbb{Z}) = \mathrm{SL}_2(\mathbb{Z})/\{\pm I\}$.) With this convention we can define the transpose of a fractional linear transformation or apply a matrix to $z \in \mathbb{H}$.

$$\gamma(z) = \frac{az+b}{cz+d} \;\Rightarrow\; \gamma^t(z) = \frac{az+c}{bz+d}, \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix} z = \frac{az+b}{cz+d}.$$

We shall employ standard notation for the identity and symplectic matrices, corresponding to identity function and involutive inversion

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad j = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

As we mentioned before, we shall let $\mathscr{L}$ denote the orbit of i under $\mathrm{SL}_2(\mathbb{Z})$. We shall employ $\mathscr{L}^*$ as an abbreviation of $\mathscr{L} - \{i\}$,

$$\mathscr{L} = \{\gamma(i) : \ \gamma \in \mathrm{SL}_2(\mathbb{Z})\},$$
$$\mathscr{L}^* = \mathscr{L} - \{i\} = \{\gamma(i) : \ \gamma \in \mathrm{SL}_2(\mathbb{Z}),\ \gamma \ne \pm I, \pm j\}.$$

Finally, we shall denote by d the hyperbolic distance corresponding to Poincaré's metric (1).

2. Preliminaries and symmetries

The group of proper motions of $\mathbb{H}$ is represented in $\mathrm{SL}_2(\mathbb{R})$ (in fact adding negative conjugation we get all motions), so

$$d\big(\gamma(z), \tau(w)\big) = d\big(z, \gamma^{-1}\tau(w)\big), \quad \forall z, w \in \mathrm{SL}_2(\mathbb{R}).$$

It is possible to write an explicit formula for the distance d (see [10] and [9] for it and its geometrical interpretation) that in the orbit of $z=i$ acquires an especially simple form (see [4])

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Rightarrow 2\cosh d(i,\gamma(i)) = a^2+b^2+c^2+d^2. \tag{2}$$

We find it convenient to state separately a calculation for further reference.

Lemma 2.1. *Let* $\gamma \in \mathrm{SL}_2(\mathbb{R})$

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Rightarrow \gamma(i) = \frac{(ac+bd)+i}{c^2+d^2} \qquad \text{and} \qquad |\gamma(i)|^2 = \frac{a^2+b^2}{c^2+d^2},$$

in particular, $(ac+bd)^2+1=(a^2+b^2)(c^2+d^2)$.

Proof. The formula for $\gamma(i)$ is just a straightforward computation, and $(ac+bd)^2+1=(a^2+b^2)(c^2+d^2)$ can be quickly obtained from $\det(\gamma\gamma^t)=1$, noting that a^2+b^2 and c^2+d^2 are the diagonal entries of $\gamma\gamma^t$ and the off-diagonal are $ac+bd$. □

It is clear that the set of Euclidean visible points has eight-fold symmetry given by the dihedral group D_4. Namely *in* $\mathbb{Z}\times\mathbb{Z}$, *one of the eight points* $(\pm x,\pm y)$, $(\pm y,\pm x)$, *is visible if and only if the rest of them are visible*. In the hyperbolic case, we have four-fold symmetry.

Lemma 2.2. *Let* $z=x+iy\in\mathscr{L}$, *then one of the points* z, $\overline{z}^{-1}$, $-z^{-1}$, $-\overline{z}$, *is visible if and only if the rest of them are visible.*

Proof. The maps $T_k\colon \mathbb{H}\longrightarrow\mathbb{H}$ given by $T_1(z)=z$, $T_2(z)=\overline{z}^{-1}$, $T_3(z)=-z^{-1}$, $T_4(z)=-\overline{z}$, are involutive isometries with $T_k(i)=i$. They leave $\mathscr{L}$ invariant because

$$T_2\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} i\right) = \begin{pmatrix} -c & d \\ -a & b \end{pmatrix} i, \qquad T_3(\gamma i) = j\gamma(i) \quad \text{and} \quad T_4 = T_2\circ T_3.$$

Hence if g is an arc of geodesic with $i\in g$ and $\#(g\cap\mathscr{L}^*)=1$, then $T_k g$ has the same property. □

This result allows to subdivide $\mathbb{H}$ in four 'quadrants' with disjoint interior:

$$Q_1=\{z\in\mathbb{H}\colon |z|\le 1,\ \mathrm{Re}(z)\ge 0\},\quad Q_2=\{z\in\mathbb{H}\colon |z|\ge 1,\ \mathrm{Re}(z)\ge 0\},$$
$$Q_3=\{z\in\mathbb{H}\colon |z|\ge 1,\ \mathrm{Re}(z)\le 0\},\quad Q_4=\{z\in\mathbb{H}\colon |z|\le 1,\ \mathrm{Re}(z)\le 0\}.$$

With the notation of the previous proof $T_k\big|_{Q_k}\colon Q_k\longrightarrow Q_1$ are well-defined isometries.

In the Euclidean case we can assign bijectively to each point in the lattice the integral translation applying the origin on it, but in the hyperbolic case the group $\mathrm{SL}_2(\mathbb{Z})$ is not faithfully represented by the orbit of i due to the fact that the stability group of $z=i$ is $\{\pm I,\pm j\}$. Hence the map

$$\begin{aligned} \mathrm{SL}_2(\mathbb{Z}) &\longrightarrow \mathscr{L}=\{\gamma(i)\colon\ \gamma\in\mathrm{SL}_2(\mathbb{Z})\} \\ \gamma &\longmapsto \gamma(i) \end{aligned}$$

is 4-to-1. We recover the Euclidean situation with some sign conventions.

Lemma 2.3. *The map*

$$\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : \begin{array}{c} a,b \geq 0,\ ac+bd>0, \\ a^2+b^2<c^2+d^2 \end{array} \right\} \longrightarrow \mathscr{L} \cap \mathrm{Int}(Q_1)$$

given by $\gamma \mapsto \gamma(i)$, *is well-defined and bijective.*

Proof. By Lemma 2.1, $ac+bd>0$ and $(a^2+b^2)/(c^2+d^2)<1$ is equivalent to $\gamma(i) \in \mathrm{Int}(Q_1)$. On the other hand, $\gamma(i) = \gamma'(i)$ if and only if $\gamma = \gamma'\tau$ with $\tau \in \{\pm I, \pm j\}$ and there is only a choice of τ giving $a,b \geq 0$. □

3. Rays and visible points

Consider a half-infinite line ℓ starting at $(0,0)$ and containing some other point of $L = \mathbb{Z} \times \mathbb{Z}$. It is fairly easy to prove that $\ell \cap L$ *is the set of non-negative multiples of a certain visible point* $P \in \ell$. On the other hand, we have the straightforward arithmetic interpretation that *visible points are simply the lattice points having coprime coordinates*. In this section we shall state the hyperbolic analog of these results.

PROPOSITION 3.1.

Let r be a ray in $\mathbb{H}$ (half-infinite geodesic) starting at i and containing some other point of $\mathscr{L}$, then there exists $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ such that

$$r \cap \mathscr{L}^* = \left\{ (\gamma\gamma^t)^{n-1}\gamma(i) :\ n \in \mathbb{Z}^+ \right\} \cup \left\{ (\gamma\gamma^t)^n(i) :\ n \in \mathbb{Z}^+ \right\}.$$

Moreover $\gamma(i)$ is visible and the points in $r \cap \mathscr{L}$ are equally spaced on r.

The proof employs the following auxiliary result:

Lemma 3.2. *Let $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ with $\gamma(i) \neq i$, then $\tau = \gamma\gamma^t$ leaves invariant the geodesic connecting i and $\gamma(i)$, and it holds $d(i,\gamma(i)) = d(\gamma(i),\tau(i))$.*

Proof. A (non-vertical infinite) geodesic g can be represented as a Euclidean semicircle in $\mathbb{H}$ orthogonal to the real axis. If $i \in g$, by simple trigonometry, end-points are $-x_0$ and x_0^{-1} for some $x_0 \in \mathbb{R}$.

Take the geodesic passing through i and $\gamma(i)$ as g, then $i \in \gamma^{-1}g$. Hence we have $\gamma^{-1}(-x_0) = -y_0$ and $\gamma^{-1}(x_0^{-1}) = y_0^{-1}$ for some $y_0 \in \mathbb{R}$. Consequently the end-points of $\gamma j^{-1}\gamma^{-1} jg$ are $\gamma j^{-1}\gamma^{-1} j(-x_0) = -x_0$ and $\gamma j^{-1}\gamma^{-1} j(x_0^{-1}) = x_0^{-1}$. It means that $\gamma j^{-1}\gamma^{-1} j$ leaves g invariant. A calculation proves $j^{-1}\gamma^{-1} j = \gamma^t$.

On the other hand, $d(i,\gamma^t(i)) = d(\gamma(i),\tau(i))$ because γ is an isometry, and $d(i,\gamma(i)) = d(i,\gamma^t(i))$ by (2). □

Proof of Proposition 3.1. Parametrizing r by arc length, each point on r is determined by its distance to i. Plainly r contains exactly one visible point, say $\gamma(i)$ and write $l = d(i,\gamma(i))$.

Note that the signs of $\mathrm{Re}(\gamma(i))$ and $\mathrm{Re}(\gamma\gamma^t(i))$ coincide (see Lemma 2.1). Then by the previous lemma $\gamma\gamma^t$ applies the half geodesic r into itself and

$$\begin{aligned} d(i,\gamma\gamma^t(i)) &= d(i,\gamma(i)) + d(\gamma(i),\tau(i)) \\ &= 2d(i,\gamma(i)) = 2l. \end{aligned}$$

Hence $\gamma\gamma^t\big|_r$ is just a translation of lenght $2l$ along r and

$$\left\{(\gamma\gamma^t)^{n-1}\gamma(i):\ n\in\mathbb{Z}^+\right\}\cup\left\{(\gamma\gamma^t)^n(i):\ n\in\mathbb{Z}^+\right\}$$
$$=\{z\in r:\ d(i,z)\in l\mathbb{Z}^+\}.$$

This set of l-spaced points is contained in $r\cap\mathscr{L}^*$. It remains to prove that any $w\in r\cap\mathscr{L}^*$ satisfies $d(i,w)\in l\mathbb{Z}^+$. If $d(i,w)\notin l\mathbb{Z}^+$ then for some $k\in\mathbb{Z}$,

$$2kl<d(i,w)<(2k+1)l \quad \text{or} \quad (2k-1)l<d(i,w)<2kl.$$

In the first case $z=(\gamma\gamma^t)^{-k}(w)\in r\cap\mathscr{L}^*$ verifies $0\neq d(i,z)<l=d(i,\gamma(i))$ which is a contradiction because $\gamma(i)$ is visible. In the second case the same argument applies with $z=j(\gamma\gamma^t)^{-k}(w)$. In this connection note that $(\gamma\gamma^t)^{-k}(w)\in r'$ where r' is the complementary ray of r, i.e. $r\cap r'=\{i\}$ and $r\cup r'$ form an infinite geodesic g; and j applies r' isometrically into r leaving i fixed (j permutes the end-points of g), then $d(i,z)=d(i,(\gamma\gamma^t)^{-k}(w))=2kl-d(i,w)$. $\square$

Now we are going to characterize visible points in terms of their coordinates. Recall firstly that any $z\in\mathscr{L}^*$ is uniquely written as

$$z=\frac{B+i}{D} \qquad \text{with } B,D\in\mathbb{Z},$$

and consider the map

$$\begin{aligned} V:\mathscr{L}^* &\longrightarrow L^*=\mathbb{Z}\times\mathbb{Z}-\{(0,0)\} \\ z=(B+i)/D &\longrightarrow (B,D-A) \end{aligned} \qquad \text{with} \quad A=(B^2+1)/D.$$

Note that it is well-defined and applies the first quadrant $\mathscr{L}^*\cap Q_1$ into the Euclidean first quadrant (see §2, esp. Lemma 2.1). By symmetry, it is enough to state visibility criterion in Q_1; in the rest of the quadrants it is similar up to sign changes.

Theorem 3.3. *Let $z\in\mathscr{L}^*\cap Q_1$ and A,B,D as before, then z is not visible if and only if there exists integers $1\le a\le b<d$ with $ad=b^2+1$ and $b|B$, such that*

$$\frac{B}{b}=\frac{D-A}{d-a}\neq 1.$$

Proof. Firstly note that (see Lemma 2.1)

$$z=\gamma(i)\ \Rightarrow\ \gamma\gamma^t=\begin{pmatrix}A & B\\ B & D\end{pmatrix}.$$

By Lemma 3.2, the hyperbolic motion $\gamma\gamma^t$ leaves invariant the geodesic g connecting i and $\gamma(i)$. End-points of g, say $z_1,z_2\in\mathbb{R}$, are the roots of the quadratic equation $(Az+B)/(Bz+D)=z$, then the Euclidean center of the semicircle representing g is

$$\frac{z_1+z_2}{2}=\frac{A-D}{2B}. \tag{3}$$

Now we shall consider both implications separately.

($\Rightarrow$) If $z=\gamma(i)$ is not visible, let $\tau(i)\neq z$ be the visible point in the ray connecting i and z. Let us take γ and τ normalised as in Lemma 2.3. By eq. (3) and Lemma 3.2

$$\tau\tau^t=\begin{pmatrix}a & b\\ b & d\end{pmatrix}\ \Rightarrow\ \frac{a-d}{2b}=\frac{A-D}{2B}$$

because $\tau(i)$ and $\gamma(i)$ are on the same ray. Of course $\tau\tau^t\in \mathrm{SL}_2(\mathbb{Z})\Rightarrow ad=b^2+1$ and it only remains to prove $b|B$, $b\neq B$. By Proposition 3.1, $\gamma=(\tau\tau^t)^n\tau$ or $\gamma=(\tau\tau^t)^n$ with $n\in\mathbb{Z}^+$. In any case $\gamma\gamma^t=(\tau\tau^t)^k$, $k\geq 2$ and by induction on k it follows that b divides the second entry of $(\tau\tau^t)^k$. The positivity given by Lemma 2.3 assures $b<B$.

($\Leftarrow$) As $ad=b^2+1$, by the theory of binary quadratic forms (see Art. 183, [7]), we can find $\tau\in\mathrm{SL}_2(\mathbb{Z})$ such that

$$\tau\tau^t=\begin{pmatrix}a & b\\ b & d\end{pmatrix}.$$

In fact we can assume that τ is as in Lemma 2.3. The relation $B/b=(D-A)/(d-a)$ and (3) imply that $\tau(i)$ and $\gamma(i)$ are in the same geodesic ray starting at i. Since $B>b$ and $D-A>d-a>0$,

$$\left.\begin{array}{r}D^2+A^2-2AD>d^2+a^2-2ad\\ 4(B^2+1)>4(b^2+1)\end{array}\right\}\Rightarrow\ (D+A)^2>(d+a)^2$$
$$\Rightarrow\ D+A>d+a$$

just adding both inequalities. Hence

$$D-A>d-a,\quad D+A>d+a\ \Rightarrow\ D>d\ \Rightarrow\ \mathrm{Im}(\gamma(i))<\mathrm{Im}(\tau(i)).$$

As $\gamma(i)$ and $\tau(i)$ belong to the same ray, the latter condition implies

$$d(i,\gamma(i))>d(i,\tau(i)),$$

thus $\gamma(i)$ is not visible. □

Example 1. Consider

$$z=\begin{pmatrix}2 & 1\\ 3 & 2\end{pmatrix}i=\frac{8+i}{13}\ \Rightarrow\ B=8,\ D=13,\ A=5,\ D-A=8.$$

In this case, the conditions of the theorem read $b=1,4$, $ad=5,17$, respectively, with $b=a-d\neq 8$. This is fulfilled for $b=1$, $d=5$, $a=1$, and hence $z=(8+i)/13$ is not visible.

Example 2. The point

$$z=\begin{pmatrix}1 & 3\\ 2 & 7\end{pmatrix}i=\frac{23+i}{53}$$

is visible because $B=23$, $D=53$, $A=10$ and as B and $D-A$ are coprime the equation $B/b=(D-A)/(d-a)\neq 1$ cannot hold.

Let us state separately the last remark:

COROLLARY 3.4.

If B and $D-A$ are coprime then z is visible. Equivalently, if $V(z)$ is visible (in Euclidean sense) then z is visible (in hyperbolic sense).

Remark. The reciprocal is not true. The simplest counterexample is $z=(2+i)/5$ which is visible with $V(z)=(B,D-A)=(2,4)$.

In the Euclidean case, if we enumerate the points on each ray starting by zero (assigned to the origin), then the points labelled by even numbers, say the *points in even place*, are exactly the sublattice of points with even coordinates. In the hyperbolic case we can define equally points in even place and the following result allows a coordinate characterization.

PROPOSITION 3.5.

$z\in\mathscr{L}$ is a point in even place if and only if $z=\tau(i)$ with τ symmetric and equivalently, if and only if

$$z=\frac{(a+d)b+i}{b^2+d^2}$$

for some integers $ad=b^2+1$.

Proof. By Proposition 3.1 we have that the points in $r\cap\mathscr{L}$, where r is the ray connecting i and z, are equally spaced. In fact we have proved that $\gamma\gamma^t\big|_r$ is a translation of length $2l$ where l is the separation between consecutive points. Hence z is in even place if and only if $\tau=(\gamma\gamma^t)^n$ where n is a non-negative integer. Then if z is in even place, τ is symmetric. Reciprocally, if τ is symmetric we can write (Art. 183, [7]) (as in the previous proof) $\tau=\delta\delta^t$, and by Proposition 3.1, $\delta=(\gamma\gamma^t)^{n-1}\gamma$ or $\delta=(\gamma\gamma^t)^n$. In any case, $\tau=(\gamma\gamma^t)^m$ and z is in even place. □

4. The visible lattice point problem

The asymptotics of the number of visible points in a Euclidean circle of large radius R has been studied by several authors. This number is usually approximated by a formula like

$$E^*(R)=\frac{6}{\pi^2}R^2+O(R^\alpha). \tag{4}$$

Surprisingly, any improvement on the trivial exponent $\alpha=1$ (see [12]) lead to considerations on Riemann Hypothesis and there are no unconditional results with $\alpha<1$. Several authors have proved (4) for some α assuming Riemann Hypothesis (using the arguments in [12] and intricate exponential sums estimates). The best conditional result so far is $\alpha=221/608+\varepsilon$ for every $\varepsilon>0$ [16].

In hyperbolic setting, the relation with Riemann Hypothesis disappears, roughly speaking because most of the lattice points stay close to the boundary and hence comparatively few points are eclipsed, and the contribution of invisible points is absorbed by error term. Considering firstly all the points in the orbit of i, visible and invisible, the asymptotics of the number of points in a large circle of radius R is given as (see [13], we introduce a $1/4$ extra factor because 4 is the cardinality of stability group of i)

$$H(R) = \frac{3}{2}\mathrm{e}^R + O(\mathrm{e}^{\alpha R}). \tag{5}$$

Using harmonic analysis on $\mathbb{H}\backslash\mathrm{SL}_2(\mathbb{Z})$ one can get $\alpha = 2/3$ (see [10], §12). This exponent has not been improved but the natural conjecture (supported by average results [4]) is $\alpha = 1/2+\varepsilon$ for every $\varepsilon > 0$.

PROPOSITION 4.1.

Let $H^*(R)$ be the number of visible points in the circle $\{z \in \mathbb{H} \colon\ d(i,z) \le R\}$, then

$$H^*(R) = H(R) - \frac{3}{2}\mathrm{e}^{R/2} + O(\mathrm{e}^{R/3}).$$

Proof. Given a ray r connecting i and some point in $\mathscr{L}^*$, let

$$r(R) = \#\{z \in \mathscr{L}^* \cap r \colon\ d(i,z) \le R\}.$$

According to Proposition 3.1, the points in $\mathscr{L}\cap r$ are l-spaced, hence

$$r(R) = \left[\frac{R}{l}\right] \qquad \text{and} \qquad 1 = \sum_{n\le R} \mu(n) r(R/n) \quad \text{for } R \ge l,$$

(see [6]) where $[\,\cdot\,]$ denotes integral part and μ is Möbius function.

Let $\mathscr{R}$ be the set of rays as before containing some $z \in \mathscr{L}^*$ with $d(i,z) \le R$. Each ray in $\mathscr{R}$ contains exactly a visible point and we have for $R > 1$,

$$H^*(R) = \sum_{r\in\mathscr{R}} 1 = \sum_{n\le R} \mu(n) \sum_{r\in\mathscr{R}} r(R/n)$$
$$= \sum_{n\le R} \mu(n) H(R/n).$$

Using (5)

$$H^*(R) = H(R) - H(R/2) - H(R/3) + O(\mathrm{e}^{\alpha R/5})$$

and taking $\alpha = 2/3$, $H(R/2) - H(R/3) = 3\mathrm{e}^{R/2}/2 + O(\mathrm{e}^{R/3})$ we get the result. Note that under the conjecture $\alpha = 1/2+\varepsilon$ we could diminish error term to $O(\mathrm{e}^{(1+\varepsilon)R/4})$ extracting an extra $-3\mathrm{e}^{R/3}/2$ term. □

The last proposition allows to translate to $H^*(R)$ some results known for $H(R)$. Following [13], we define the normalized remainder

$$\Delta^*(R) = \frac{H^*(R) - 3\mathrm{e}^R/2}{\mathrm{e}^{R/2}}.$$

It turns out that $\Delta^*(R)$ is biased (because of the influence on invisible points), and inherits the oscillation of $H(R)$. After Proposition 4.1, this is just a consequence of the main results of [13].

COROLLARY 4.2.

The mean value of $\Delta^*(R)$ *is* $3/2$, *i.e.*

$$\lim_{R\to\infty} \frac{1}{R}\int_1^R \Delta^*(t)\,\mathrm{d}t = 3/2,$$

but $\Delta^*(R)$ *is not bounded. In fact*

$$\limsup_{R\to\infty} \frac{\Delta^*(R)}{(\log R)^\delta} = \infty \qquad \textit{for every} \quad \delta < 1/4.$$

Proof. Let $\Delta(R) = (H(R) - 3\mathrm{e}^R/2)/\mathrm{e}^{R/2}$. By Theorems 1.1 and 1.2 in [13] (note that for the full modular group $E(z,s) = \zeta_Q(s)/\zeta(2s)$ where ζ_Q is an Epstein zeta function [9]), it holds that

$$\lim_{R\to\infty} \frac{1}{R}\int_1^R \Delta(t)\,\mathrm{d}t = 0 \quad \text{and} \quad \limsup_{R\to\infty} \frac{\Delta(R)}{(\log R)^\delta} = \infty.$$

By Proposition 4.1, $\Delta^*(R) = \Delta(R) - 3/2 + O(\mathrm{e}^{-R/6})$ and the claimed results follow. □

We have also some control on a quantity related to the variance.

COROLLARY 4.3.

$$\limsup_{R\to\infty} \frac{1}{R}\int_1^R |t^{-1}\Delta^*(t)|^2\,\mathrm{d}t < \infty.$$

Proof. Corollary 2.1.1 of [4] implies

$$\int_X^{2X} \left| H\left(\text{arc cosh}\,\frac{x}{2}\right) - \frac{3}{2}x \right|^2 \mathrm{d}x = O(X^2\log^2 X).$$

With the change of variables $x = 2\cosh t$ and writing $r = \log X$, we have

$$\int_r^{r+1} \left| H(t) - \frac{3}{2}\mathrm{e}^t \right|^2 \mathrm{e}^t\,\mathrm{d}t = O(r^2\mathrm{e}^{2r}).$$

Using Proposition 4.1, and after some manipulations we get that

$$\int_r^{r+1} \left| \frac{\Delta^*(t)}{r} \right|^2 \mathrm{d}t$$

is bounded. Summing on $1 \le r \le R-1$, the result is proved. □

5. The orchard problem

It is known that the solution of the orchard problem (as stated in the Introduction) is that the view is obstructed in a circular orchard of radius R if and only if the trunks have radii $\varepsilon \ge R^{-1} + f(R)$ for certain $f(R) = O(R^{-2})$ (an 'exact' formula is given in [2]). One can also consider, so to speak, negative orchard problem, asking for the maximal ε such that it is possible to see all the visible points in the circle of radius R. In Euclidean setting the solution is the same as that of the original problem, but in hyperbolic setting both problems considerably differ.

We shall associate to each $z \in \mathscr{L}^*$ and $\varepsilon > 0$ the thick point z_ε with radius ε; this means the circle $z_\varepsilon = \{w \in \mathbb{H}\colon\ d(w,z) \le \varepsilon\}$.

DEFINITION

Given $z, w \in \mathscr{L}^*$ we say that z_ε *eclipses* w if $r \cap z_\varepsilon \neq \emptyset$ where r is the geodesic arc connecting i and w.

Our main tool for treating the obstruction of view in $\mathscr{L}$ is the following result.

PROPOSITION 5.1.

Let $z, w \in \mathscr{L}^* \cap Q_1$ with $d(i,z) \le d(i,w)$, say $z = \gamma(i)$ and $w = \tau(i)$. Then z_ε eclipses w if and only if

$$\sinh \varepsilon \ge \frac{\left|\mathrm{Tr}(\gamma\gamma^t j\tau\tau^t)\right|}{2\sinh d(i,w)},$$

where Tr *indicates the trace.*

Proof. Note firstly that z_ε eclipses w if and only if ε is greater than the distance of z to the geodesic g connecting i and w, because the foot of the perpendicular through z, say F, belongs to Q_1 and $d(i,F) \le d(i,z) \le d(i,w)$ (by hyperbolic Pythagorean theorem $\cosh a \cosh b = \cosh c$).

Whence we are going to prove that for every $\gamma, \tau \in \mathrm{SL}_2(\mathbb{R})$, $z = \gamma(i)$, $w = \tau(i) \neq i$, if g is the (infinite) geodesic through i and w, we have

$$\sinh d(z,g) = \frac{\left|\mathrm{Tr}(\gamma\gamma^t j\tau\tau^t)\right|}{2\sinh d(i,w)}. \tag{6}$$

Consider $m_\theta \in \mathrm{SL}_2(\mathbb{R})$ given by

$$m_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

In hyperbolic plane this is a rotation at i of angle 2θ (§1.2, [10]). Both sides in (6) are invariant under the changes $\gamma \mapsto m_\theta \gamma m_{\theta'}$ and $\tau \mapsto m_\theta \tau m_{\theta'}$ because $m_\theta, m_{\theta'}$ are orthogonal matrices, leave i invariant and $m_\theta^t j m_\theta = j$. With a suitable choice of m_θ and $m_{\theta'}$ we can assume by Cartan's decomposition (§1.3, [10]) that

$$\tau = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} \quad \text{and} \quad z = \lambda^2 i \quad \text{for some } \lambda \in \mathbb{R}^+.$$

Let $\gamma(u) = (au+b)/(cu+d)$, then $\left|\mathrm{Tr}(\gamma\gamma^t j\tau\tau^t)\right| = |ac+bd|(\lambda^2 - \lambda^{-2})$ and using (2), we have that (6) reduces to prove that the hyperbolic distance D from $z = \gamma(i)$ to the imaginary axis is given by $\sinh D = |ac+bd|$. As z is in the circle $|\zeta| = |\gamma(i)|$ which is orthogonal to this axis, by hyperbolic Pythagorean theorem

$$\sinh^2 D = \frac{\cosh^2 d(i,\gamma(i))}{\cosh^2 d(i,|\gamma(i)|i)} - 1,$$

which using (2), Lemma 2.1 and (1), gives the result. $\square$

First let us consider the negative orchard problem.

PROPOSITION 5.2.

Let $C_R^* = \mathscr{L}^* \cap \{z\colon\ d(i,z) \le R\}$. If $\varepsilon < 2\mathrm{e}^{-R}$ then none of the points in C_R^* enlarged to radius ε eclipses another point in C_R^*.

Proof. Let $z, w \in C_R^* \cap Q_1$, say $d(i,z) \le d(i,w)$, and let d_1 and d_2 be the distances of z and w to the geodesics connecting i with w and i with z, respectively. Using sine rule [9]

$$\frac{\sinh d_1}{\sinh d(i,z)} = \frac{\sinh d_2}{\sinh d(i,w)},$$

hence $d_1 \le d_2$ and w_ε eclipses z implies that z_ε eclipses w.

It is easy to check that for a symmetric matrix in $\mathrm{SL}_2(\mathbb{Z})$ the off-diagonal entry and the trace are congruent modulo 2. If B and β are the off-diagonal entries of $\gamma\gamma^t$ and $\tau\tau^t$, a calculation proves

$$\begin{aligned}\mathrm{Tr}(\gamma\gamma^t j\tau\tau^t) &\equiv B\mathrm{Tr}(\tau\tau^t) + \beta\mathrm{Tr}(\gamma\gamma^t) \pmod 2,\\ &\equiv 2\mathrm{Tr}(\tau\tau^t)\mathrm{Tr}(\gamma\gamma^t) \equiv 0 \pmod 2.\end{aligned}$$

By Proposition 5.1, if z_ε eclipses w then $\sinh\varepsilon \ge (\sinh R)^{-1}$, and this implies $\varepsilon \ge 2\mathrm{e}^{-R}$.

If against our assumption z and w do not belong to Q_1 but the corresponding rays determine an acute angle, then the same proof applies after a suitable rotation. If the angle is not acute, if z_ε eclipses to w then $i \in z_\varepsilon$ and, according to (2), $\cosh\varepsilon \ge 3/2$, i.e. $\varepsilon \ge 0.9624\ldots$ and we can assume $2\mathrm{e}^{-R} < 0.764\ldots$ because otherwise $C_R^* = \emptyset$. □

A construction using the properties of Fibonacci numbers allows to show that the previous result is sharp.

PROPOSITION 5.3.

Given $C > 2$ there exist sequences of values $z \in \mathbb{H}$, $w \in \mathbb{H}$ and $R \in \mathbb{R}$ tending to ∞ such that $z, w \in C_R^*$ are visible points and z_ε eclipses w with $\varepsilon = C\mathrm{e}^{-R}$.

Proof. Consider Fibonacci sequence $\{F_n\}_{n=1}^\infty = \{1,1,2,3,5,8,\ldots\}$ and for each n,

$$\gamma = \begin{pmatrix} F_{6n-1} & F_{6n-2} \\ F_{6n+1} & F_{6n} \end{pmatrix} \quad \text{and} \quad \tau = \begin{pmatrix} F_{6n-1} & F_{6n} \\ F_{6n+1} & F_{6n+2} \end{pmatrix}.$$

It holds that $\gamma, \tau \in \mathrm{SL}_2(\mathbb{Z})$ (use that F_{k+1}/F_k are the convergents of the golden ratio, or employ the recurrence formula, 4.2.3(d) and 4.3.9(b) of [11]). Choose $z = \gamma(i)$, $w = \tau(i)$ and $R = d(i,w)$. By Lemma 2.3, $z, w \in \mathscr{L}^* \cap Q_1$ and we are under the hypothesis of Proposition 5.1.

Using the properties of Fibonacci numbers (see (4.2) and 4.2.3(d) of [11]) we get

$$\gamma\gamma^t = \begin{pmatrix} F_{12n-3} & F_{12n-1} \\ F_{12n-1} & F_{12n+1} \end{pmatrix} \quad \text{and} \quad \tau\tau^t = \begin{pmatrix} F_{12n-1} & F_{12n+1} \\ F_{12n+1} & F_{12n+3} \end{pmatrix}.$$

Take $m = 12n-3$ or $12n-1$. By Euclidean algorithm F_{m+1} and F_m are coprime, and $F_{m+1}+F_m$ and $F_{m+1}-F_m$ are coprime too (both are odd numbers). Then $F_{m+2} = F_{m+1} + F_m$ and $F_{m+4} - F_m = 2F_{m+2} + F_{m+1} - F_m$ are also coprime. By Corollary 3.4, z and w are visible.

A calculation shows

$$\begin{aligned}\mathrm{Tr}(\gamma\gamma^t j\tau\tau^t) &= F_{12n-1}(F_{12n-1} - F_{12n+3}) + F_{12n+1}(F_{12n+1} - F_{12n-3})\\ &= -F_{12n-1}(F_{12n} + F_{12n+2}) + F_{12n+1}(F_{12n-2} + F_{12n})\\ &= (F_{12n+1}F_{12n-2} - F_{12n-1}F_{12n})\\ &\quad - (F_{12n+2}F_{12n-1} - F_{12n}F_{12n+1})\\ &= -1-1 = -2.\end{aligned}$$

Where we have firstly used that $F_{k+2}-F_{k-2}=F_{k+1}+F_{k-1}$ and secondly, as before, that F_{k+1}/F_k are the convergents of $(1+\sqrt{5})/2$.

By Proposition 5.1 we have that for $\sinh\varepsilon \ge (\sinh R)^{-1}$, z_ε eclipses w, and this inequality holds with $\varepsilon = C\mathrm{e}^{-R}$ for large enough R. $\square$

It turns out (see the proof below) that the unique way of blocking completely the view from the origin in a hyperbolic orchard is enlarging a certain fixed quantity the trunks of the first four trees. In this sense, orchard problem becomes trivial in its original form.

PROPOSITION 5.4.

Every w with $d(i,w)>R$ is eclipsed by some z_ε with $z\in C_R^*$ if and only if $\varepsilon \ge \log(1+\sqrt{2})$.

Proof. If $z=\gamma(i)=(ai+b)/(ci+d)$, $z\ne i$, we have shown at the end of the proof of Proposition 5.1 that the distance D from z to the imaginary axis verifies $\sinh D = |ac+bd|$, hence $\sinh D\ge 1$ and we cannot block the view along the imaginary axis if $\varepsilon < \operatorname{arc}\sinh 1 = \log(1+\sqrt{2})$.

On the other hand, let $z_2 = 1+i\in Q_2\cap\mathscr{L}^*$. The circle $\{w:\ d(z_2,w)\le \log(1+\sqrt{2})\}$ correspond to the Euclidean circle in $\mathbb{H}$ given by $(x-1)^2+(y-\sqrt{2})^2\le 1$ (see §1.1, [10]). Applying $T_k^{-1}T_2$ we get four intersecting circles around the origin blocking the view from $z=i$. $\square$

Even disregarding near points, if we argue heuristically thinking that the points in C_R^* are uniformly distributed along the boundary (of length $2\pi\sinh R$), we can expect maximal spacing as large as $\mathrm{e}^{R/2}$ (in particular unbounded, in contrast with the Euclidean case). This effectively happens when we pass from a quadrant to another. For instance, the rays r_-, r_+ connecting i and $i-n$, $i+n$ are consecutive in the circle C_R^* where $\cosh R=(n^2+2)/2$ and the spacing $d(i-n,r_+)=d(i+n,r_-)$ is comparable to $2\mathrm{e}^{R/2}$. Applying elements of $\mathrm{SL}_2(\mathbb{Z})$ the same phenomenon repeats at different scales inside each quadrant.

6. Numerical results

Using Theorem 3.3 it is easy to write a computer program distinguishing visible from invisible points in a large hyperbolic circle. The cumulative number of them is given in table 1

Table 1. .

e^R	Visible	Invisible	Error
1000	1436	60	−16.56
2000	2904	92	−28.91
3000	4408	100	−9.84
4000	5960	124	54.86
5000	7336	140	−57.93
6000	8844	148	−39.81
7000	10372	160	−2.50
8000	11792	176	−73.83
9000	13280	176	−77.69
10000	14880	184	30.00

Table 2. .

e^R	Invisible	Approx.
1000	60	63.66
2000	92	87.52
3000	100	105.53
4000	124	120.58
5000	140	133.75
6000	148	145.60
7000	160	156.45
8000	176	166.51
9000	176	175.93
10000	184	184.82

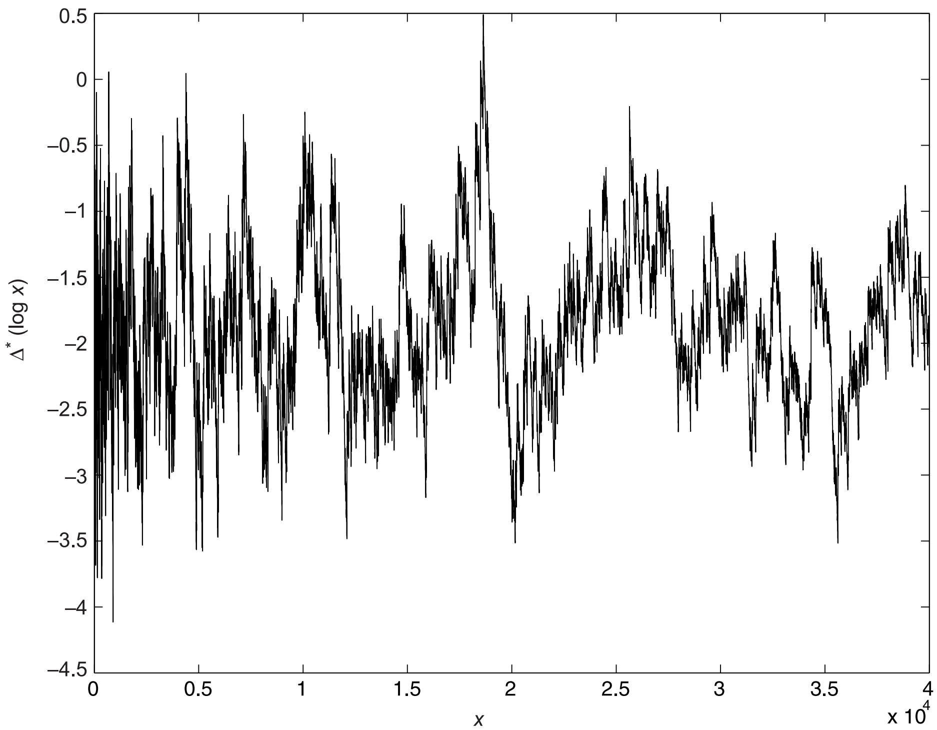

Figure 1.

The error is given by the O-term in Proposition 4.1 after approximating $H(R)$ by $3e^R/2$, i.e.,

$$\text{Error} = \text{visible} - \frac{3}{2}e^R + \frac{3}{2}e^{R/2}.$$

Note that the number of invisible points is relatively small, in accordance with Proposition 4.1. In fact, following the arguments of its proof and truncating the series $\sum \mu(n)H(R/n)$ to $n \leq 6$, one can expect

$$\text{Approx.} = \frac{3}{2}(e^{R/2} + e^{R/3} + e^{R/5} - e^{R/6})$$

to be a good approximation for the number of invisible points. Table 2 confirms this assertion for the previous data

Finally we show the graph of $\Delta^*(\log x)$ (we are approximating $H^*(\log x)$ by $H^*(\text{arc}\cosh(x/2))$, actually) (figure 1). Note the bias predicted by Corollary 4.2 due to invisible points.

The aspect of this graphic does not differ from the graphics of normalized error term in classical circle and divisor problem, but in this case it is not known that there is a limit distribution (cf. [8]).

Acknowledgements

I owe my deepest gratitude to E Valenti, who was fond of non-hyperbolic orchards.

References

[1] Allen T T, On the arithmetic of phase locking: Coupled neurons as a lattice on $\mathbb{R}^2$, *Phys.* **D6** (1983) 305–320
[2] Allen T T, Pólya's orchard problem, *Am. Math. Monthly* **93** (1986) 98–104
[3] Baake M, Moody R V and Pleasants P A B, Diffraction from visible lattice points and *k*th power free integers. *Discrete Math.* **221** (2000) 3–42
[4] Chamizo F, Some applications of large sieve in Riemann surfaces, *Acta Arith.* **77** (1996) 315–337
[5] Erdös P, Gruber P M and Hammer J, Lattice points. Pitman Monographs and Surveys in Pure and Applied Mathematics, 39. Longman Scientific & Technical (1989)
[6] Ellison W J, Les nombres premiers. En collaboration avec Michel Mendès France (Paris: Hermann) (1975)
[7] Gauss C F, Disquisitiones arithmeticae (New York: Springer-Verlag) (1986)
[8] Heath-Brown D R, The distribution and moments of the error term in the Dirichlet divisor problem, *Acta Arith.* **60** (1992) 389–415
[9] Huxley M N, Introduction to Kloostermania. Elementary and analytic theory of numbers, Banach Center Publ., 17 (Warsaw: PWN) (1985) pp. 217–306
[10] Iwaniec H, Introduction to the spectral theory of automorphic forms. Biblioteca de la Revista Matemática Iberoamericana. Revista Matemática Iberoamericana (1995)
[11] Lovász L, Pelikán J and Vesztergombi K, Discrete mathematics. Elementary and beyond. Undergraduate Texts in Mathematics (New York: Springer-Verlag) (2003)
[12] Nowak W G, Primitive lattice points in rational ellipses and related arithmetic functions, *Monatsh. Math.* **106** (1988) 57–63
[13] Phillips R and Rudnick Z, The circle problem in the hyperbolic plane, *J. Funct. Anal.* **121** (1994) 78–116
[14] Pólya G and Szegö G, Problems and theorems in analysis, II. Theory of functions, zeros, polynomials, determinants, number theory, geometry (New York, Heidelberg: Springer-Verlag) (1976)
[15] Wesson P S, Valle K and Stabell R, The extragalactic background light and a definitive resolution of Olbers's paradox, *Astrophys. J.* **317** (1987) 601–606
[16] Wu J, On the primitive circle problem, *Monatsh. Math.* **135** (2002) 69–81